PCT/GB 2003 / 0 0 1 5 4 3

The Patent Office

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

RECD 20 JUN 2003
WIPO PCT

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed R. Mahoney

Dated 16 APR 2003

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

1/77

05APR02 E708918-5 000239
P01/7700 0.00-0207902.8

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

THE PATENT OFFICE
B
- 5 APR 2002
NEWPORT

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference PC/JM/P12113GB

2. Patent application number *(The Patent Office will fill in this part)* 0207902.8 05 APR 2002

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

THE UNIVERSITY COURT OF THE UNIVERSITY
OF EDINBURGH
OLD COLLEGE
SOUTH BRIDGE
EDINBURGH EH8 9YL
UNITED KINGDOM

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation

8358327001

4. Title of the invention "SCHIZOPHRENIA ASSOCIATED GENE (I)"

5. Name of your agent *(if you have one)* CRUIKSHANK & FAIRWEATHER

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

19 ROYAL EXCHANGE SQUARE
GLASGOW G1 3AE
UNITED KINGDOM

Patents ADP number *(if you know it)* 547002

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	29
Claim(s)	-
Abstract	-
Drawing(s)	7 x 1

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Cruikshank & Fairweather Date 4/4/02

CRUIKSHANK & FAIRWEATHER 4 APRIL 2002

12. Name and daytime telephone number of person to contact in the United Kingdom DR. P. CHAPMAN 0141-221-5767

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

SCHIZOPHRENIA ASSOCIATED GENE (I)

The present invention relates to the identification of a gene which has been disrupted in a patient diagnosed as suffering from schizophrenia and/or bi-polar affective disorder, as well as proteins encoded by the gene and antibodies thereto and to uses of such products as medicaments for treating schizophrenia and/or affective psychosis. The invention also relates to methods for diagnosing patients suffering or predisposed to schizophrenia and/or affective psychosis, as well as screens for developing novel treatment regimes for schizophrenia and/or affective psychosis.

Schizophrenia and Bipolar Affective Disorder are common and debilitating psychiatric disorders. Despite a wealth of information on the epidemiology, neuroanatomy and pharmacology of the illness, it is uncertain what molecular pathways are involved and how impairments in these affect brain development and neuronal function. Despite an estimated heritability of 60-80%, very little is known about the number or identity of genes involved in these psychoses. Although there has been recent progress in linkage and association studies, especially from genome-wide scans, these studies have yet to progress from the identification of susceptibility loci or candidate genes to the full characterisation of disease-causing genes (Berrettini, 2000).

The cloning of breakpoints in patients with chromosome abnormalities (translocations, inversions etc.) has proved instrumental in the identification of many disease genes (e.g. Duchenne Muscular Dystrophy, Retinoblastoma, Wilm's Tumour, Familial Polyposis Coli, Fragile-X Syndrome, Polycystic Kidney Disease, many leukaemias and, very recently, a candidate speech and language disorder gene (Lai et al, 2001)). Such studies assume that the chromosomal breakpoints give rise to the clinical symptoms by either directly disrupting gene sequences or perturbing gene expression. In the same way that gene-trap mutagenesis can be used to identify disrupted mouse genes (Brennan & Skarnes, 1999), the physical "flag" created by a cytogenetic breakpoint provides a geographical pointer for the disease locus.

It is amongst the objects of the present invention to provide a gene and/or protein postulated to be involved with the development and/or symptoms associated with schizophrenia and/or affective psychosis.

As will be seen, the present invention is based on the molecular characterisation of a chromosomal disruption in a subject diagnosed as suffering from a schizo affective disorder. A high-throughput Fluorescence *in situ* Hybridisation (FISH)-based approach has been adopted to map the chromosomal breakpoints in this patient. Consultation of the sequence data at the breakpoint locus not only allows efficient FISH probe selections to be made by the targeting of coding regions, but also proof of gene

disruption can be made entirely by relating the exact position of probes to the genomic structure of a candidate gene.

The informative breakpoint in the subject lies near to a gene involved in the N-Linked Glycosylation pathway; N33. This pathway consists of three stages. Firstly the assembly of a donor oligosaccharide at the endoplasmic reticulum lumen membrane. Secondly, the transfer of this molecule onto newly translated secretory and transmembrane proteins catalyzed by the oligosaccharyltransferase (OST) complex. Thirdly, there is subsequent modification of the oligosaccharides on the glycoprotein. N33 encodes a protein thought to be involved in the second stage of the pathway by analogy with yeast homologues. Without wishing to be bound by theory it is hypothesised that the breakpoint in the subject perturbs N33 expression indirectly through position effect silencing or separation of regulatory elements from the gene promoter (both effects have been shown to occur even when the breakpoints are up to 1Mb from the target gene in some instances (Kleinjan et al 1999)).

As the N33 gene is located within a chromosomal region repeatedly found positive in schizophrenia linkage studies the present inventors pursued this gene further by association study.

Certain microsatellite repeat haplotypes have been identified at the N33 locus which are over-represented in schizophrenic patients and their families compared to the

normal population. Subsequent sequencing of the N33 gene in haplotype carrying individuals is ongoing in order to identify causative mutations.

In a first aspect the present invention provides use of a polynucleotide fragment comprising the *N33* gene or a fragment, derivative or homologue thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

In another aspect the present invention provides use of a polypeptide fragment encoded by the *N33* gene, or a fragment, derivative or homologue thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

Schizophrenia and/or affective psychosis as used herein relates to schizophrenia, as well as other affective psychoses such as those listed in "The ICD-10 Classification of Mental and Behavioural Disorders" World Health Organization, Geneva 1992. Categories F20 to F29 inclusive includes Schizophrenia, schizotypal and delusional disorders. Categories F30 to F39 inclusive are Mood (affective) disorders that include bipolar affective disorder and depressive disorder. Mental Retardation is coded F70 to F79 inclusive. The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). American Psychiatric Association, Washington DC. 1994. Include all conditions coded 295.xx (Schizophrenia and Other Psychotic Disorders) and 296.xx (Major Depressive Disorders and Bipolar Disorders). Mental Retardation is

coded 315, 317, 318 and 319.

N33 has been previously cloned and sequenced and the sequence is present in the public database (Nucleic acid sequence; U42349, Protein sequence; Q13454) and described in MacGrogan et al, 1996. The genomic sequences corresponding to this gene are also present in the public database (eg. for BAC RP11-23j14) but some SNP polymorphisms or sequencing errors (eg. an extra "C" present in exon 1b, see hereinafter - cctgcccCaccggg - may result in differences to the sequences presented herein. Nevertheless, the prior art does not suggest any link between N33 and schizophrenia and affective psychosis.

In addition to the sequences previously identified, the present inventors have identified a new start exon (1a, see Figures 1 and 2) and have observed the complexity of the exon splicing at the 3' end of the gene (see Figures 1 and 2).

Thus, references herein to the *N33* gene are understood to relate to the sequences in the public databases and identified in Figures 2 and 3 and references to the N33 protein sequence are understood to relate to the sequences in the public databases and identified in Figures 2 and 3.

In certain jurisdictions claims to methods of treatment are permissible and so the skilled reader will appreciate that the *N33* gene, or fragment, derivative or homologue thereof; or N33 protein, or functionally active fragment, derivative, or homologue thereof, may be administered to an individual as a method of treating an

individual with schizophrenia and/or affective psychosis.

"Polynucleotide fragment" as used herein refers to a chain of nucleotides such as deoxyribose nucleic acid (DNA) and transcription products thereof, such as RNA. Naturally, the skilled addressee will appreciate the whole naturally occurring human genome is not included in the definition of polynucleotide fragment.

The polynucleotide fragment can be isolated in the sense that it is substantially free of biological material with which the whole genome is normally associated *in vivo*. The isolated polynucleotide fragment may be cloned to provide a recombinant molecule comprising the polynucleotide fragment. Thus, "polynucleotide fragment includes double and single stranded DNA, RNA and polynucleotide sequences derived therefrom, for example, subsequences of said fragment and which are of any desirable length. Where a nucleic acid is single stranded then both a given strand and a sequence or reverse complementary thereto is within the scope of the present invention.

In general, the term "expression product" or "gene product" refers to both transcription and translation products of said polynucleotide fragments. When the expression or gene product is a "polypeptide" (i.e. a chain or sequence of amino acids displaying a biological activity substantially similar (eg. 98%, 95%, 90%, 80%, 75% activity) to the biological activity of the protein), it does not refer to a specific length of the product as such.

Thus, the skilled addressee will appreciate that "polypeptide" encompasses *inter alia* peptides, polypeptides and proteins. The polypeptide if required, can be modified *in vivo* and *in vitro*, for example by glycosylation, amidation, carboxylation, phosphorylation and/or post-translational cleavage.

The present invention further provides a recombinant or synthetic polypeptide for the manufacture of reagents for use as therapeutic agents in the treatment of schizophrenia and/or affective psychosis. In particular, the invention provides pharmaceutical compositions comprising the recombinant or synthetic polypeptide together with a pharmaceutically acceptable carrier therefor.

The present invention further provides an isolated polynucleotide fragment capable of specifically hybridising to a related polynucleotide sequence from another species. In this manner, the present invention provides probes and/or primers for use in *ex vivo* and/or *in situ* detection and expression studies. Typical detection studies include polymerase chain reaction (PCR) studies, hybridisation studies, or sequencing studies. In principle any specific polynucleotide sequence fragment from the identified sequences may be used in detection and/or expression studies. The skilled addressee understands that a specific fragment is a fragment of the sequence which is of sufficient length, generally greater than 10, 12, 14, 16 or 20 nucleotides in length, to bind specifically to the

sequence, under conditions of high stringency, as defined herein, and not bind to unrelated sequences, that is sequences from elsewhere in the genome of the organism other than an allelic form of the sequence or non-homologous sequences from other organisms.

"Capable of specifically hybridising" is taken to mean that said polynucleotide fragment preferably hybridises to a related or similar polynucleotide sequence in preference to unrelated or dissimilar polynucleotide sequences.

The invention includes polynucleotide sequence(s) which are capable of specifically hybridising to an polynucleotide fragment as described herein or to a part thereof without necessarily being completely complementary or reverse complementary to said related polynucleotide sequence or fragment thereof. For example, there may be at least 50%, or at least 75%, at least 90%, or at least 95% complementarity. Of course, in some cases the sequences may be exactly reverse complementary (100% reverse complementary) or nearly so (e.g. there may be less than 10, typically less than 5 mismatches). Thus, the present invention also provides anti-sense or complementary nucleotide sequence(s) which is/are capable of specifically hybridising to the disclosed polynucleotide sequence. If a specific polynucleotide is to be used as a primer in PCR and/or sequencing studies, the polynucleotide must be capable of hybridising to related nucleic acid and capable of initiating chain extension from 3' end of the polynucleotide, but not able to correctly initiate chain

extension from unrelated sequences.

If a polynucleotide sequence of the present invention is to be used in hybridisation studies to obtain or identify a related sequence from another organism the polynucleotide sequence should preferably remain hybridised to a sample polynucleotide under stringent conditions. If desired, either the test or sample polynucleotide may be immobilised. Generally the test polynucleotide sequence is at least 10, 14, 20 or at least 50 bases in length. It may be labelled by suitable techniques known in the art. Preferably the test polynucleotide sequence is at least 200 bases in length and may even be several kilobases in length. Thus, either a denatured sample or test sequence can be first bound to a support. Hybridization can be effected at a temperature of between 50 and 70°C in double strength SSC (2xNaCl 17.5g/l and sodium citrate (SC) at 8.8g/l) buffered saline containing 0.1% sodium dodecyl sulphate (SDS). This can be followed by rinsing of the support at the same temperature but with a buffer having a reduced SSC concentration. Depending upon the degree of stringency required, and thus the degree of similarity of the sequences, such reduced concentration buffers are typically single strength SSC containing 0.1%SDS, half strength SSC containing 0.1%SDS and one tenth strength SSC containing 0.1%SDS. Sequences having the greatest degree of similarity are those the hybridisation of which is least affected by washing in buffers of reduced concentration. It is most preferred that the sample and inventive

sequences are so similar that the hybridisation between them is substantially unaffected by washing or incubation in standard sodium citrate (0.1 x SSC) buffer containing 0.1%SDS.

Oligonucleotides may be designed to specifically hybridise to N33 nucleic acid. They may be synthesised, by known techniques and used as primers in PCR or sequencing reactions or as probes in hybridisations designed to detect the presence of a mutated or normal N33 gene in a sample. The oligonucleotides may be labelled by suitable labels known in the art, such as, radioactive labels, chemiluminescent labels or fluorescent labels and the like.

The term "oligonucleotide" is not meant to indicate any particular length of sequence and encompasses nucleotides of preferably at least 10b (e.g. 10b to 1kb) in length, more preferably 12b-500b in length and most preferably 15b to 100b.

The oligonucleotides may be designed with respect to any of the sequences described herein and may be manufactured according to known techniques. They may have substantial sequence identity (e.g. at least 50%, at least 75%, at least 90% or at least 95% sequence identity) with one of the strands shown therein or an RNA equivalent, or with a part of such a strand. Preferably such a part is at least 10, at least 30, at least 50 or at least 200 bases long. It may be an open reading frame (ORF) or a part thereof.

Oligonucleotides which are generally greater than 30 bases in length should preferably remain hybridised to a sample polynucleotide under one or more of the stringent conditions mentioned above. Oligonucleotides which are generally less than 30 bases in length should also preferably remain hybridised to a sample polynucleotide but under different conditions of high stringency. Typically the melting temperature of an oligonucleotide less than 30 bases may be calculated according to the formula of; 2°C for every A or T, plus 4°C for every G or C, minus 5°C. Hybridization may take place at or around the calculated melting temperature for any particular oligonucleotide, in 6 x SSC and 1% SDS. Non specifically hybridised oligonucleotides may then be removed by stringent washing, for example in 3 x SSC and 0.1% SDS at the same temperature. Only substantially similar matched sequences remain hybridised i.e. said oligonucleotide and corresponding test nucleic acid.

When oligonucleotides of generally less than 30 bases in length are used in sequencing and/or PCR studies, the melting temperature may be calculated in the same manner as described above. The oligonucleotide may then be allowed to anneal or hybridise at a temperature around the oligonucleotides calculated melting temperature. In the case of PCR studies the annealing temperature should be around the lower of the calculated melting temperatures for the two priming oligonucleotides. It is to be appreciated that the conditions and melting temperature calculations

are provided by way of example only and are not intended to be limiting. It is possible through the experience of the experimenter to vary the conditions of hybridisation and thus anneal/hybridise oligonucleotides at temperatures above their calculated melting temperature. Indeed this can be desirable in preventing so-called non-specific hybridisation from occurring.

It is possible when conducting PCR studies to predict an expected size or sizes of PCR product(s) obtainable using an appropriate combination of two or more oligonucleotides, based on where they would hybridise to the sequences described herein. If, on conducting such a PCR on a sample of DNA, a fragment of the predicted size is obtained, then this is predictive that the DNA encodes a homologous sequence from a test organism.

Proteins for all the applications described herein can be produced by cloning the gene for example into plasmid vectors that allow high expression in a system of choice e.g. insect cell culture, yeast, animal cells, bacteria such as *Escherichia coli*. To enable effective purification of the protein, a vector may be used that incorporates an epitope tag (or other "sticky" extension such as His6) onto the protein on synthesis. A number of such vectors and purification systems are commercially available.

The polynucleotide fragment can be molecularly cloned into a prokaryotic or eukaryotic expression vector using standard techniques and administered to a host. The expression vector is taken up by cells and the

polynucleotide fragment of interest expressed, producing protein.

It will be understood that for the particular polypeptides embraced herein, natural variations such as may occur due to polymorphisms, can exist between individuals or between members of the family. These variations may be demonstrated by (an) amino acid difference(s) in the overall sequence or by deletions, substitutions, insertions, inversions or additions of (an) amino acid(s) in said sequence. All such derivatives showing the recognised activity are included within the scope of the invention. For example, for the purpose of the present invention conservative replacements may be made between amino acids within the following groups:

(I) Alanine, serine, threonine;

(II) Glutamic acid and aspartic acid;

(III) Arginine and leucine;

(IV) Asparagine and glutamine;

(V) Isoleucine, leucine and valine;

(VI) Phenylalanine, tyrosine and tryptophan

Moreover, recombinant DNA technology may be used to prepare nucleic acid sequences encoding the various derivatives outlined above.

As is well known in the art, the degeneracy of the genetic code permits substitution of bases in a codon resulting in a different codon which is still capable of coding for the same amino acid, e.g. the codon for amino acid glutamic acid is both GAT and GAA. Consequently, it

is clear that for the expression of polypeptides from nucleotide sequences described herein or fragments thereof, use can be made of a derivative nucleic acid sequence with such an alternative codon composition different from the nucleic acid sequences shown in the Figures.

The polynucleotide fragments of the present invention are preferably linked to regulatory control sequences. Such control sequences may comprise promoters, operators, inducers, enhancers, silencers, ribosome binding sites, terminators etc. Suitable control sequences for a given host may be selected by those of ordinary skill in the art.

A polynucleotide fragment according to the present invention can be ligated to various expression controlling sequences, resulting in a so called recombinant nucleic acid molecule. Thus, the present invention also includes an expression vector containing an expressible nucleic acid molecule. The recombinant nucleic acid molecule can then be used for the transformation of a suitable host.

Specific vectors which can be used to clone nucleic acid sequences according to the invention are known in the art (e.g. Rodriguez, R.L. and Denhadt, D.T., Edit., Vectors: a survey of molecular cloning vectors and their uses, Butterworths, 1988, or Jones *et al.*,Vectors: Cloning Applications: Essential Techniques (Essential techniques series), John Wiley & Son. 1998).

The methods to be used for the construction of a recombinant nucleic acid molecule according to the invention are known to those of ordinary skill in the art

and are inter alia set forth in Sambrook, et al. (Molecular Cloning: a laboratory manual Cold Spring Harbour Laboratory, 1989).

The present invention also relates to a transformed cell containing the polynucleotide fragment in an expressible form. "Transformation", as used herein, refers to the introduction of a heterologous polynucleotide fragment into a host cell. The method used may be any known in the art, for example, direct uptake, transfection transduction or electroporation (Current Protocols in Molecular Biology, 1995. John Wiley and Sons Inc.). The heterologous polynucleotide fragment may be maintained through autonomous replication or alternatively, may be integrated into the host genome. The recombinant nucleic acid molecules preferably are provided with appropriate control sequences compatible with the designated host which can regulate the expression of the inserted polynucleotide fragment, e.g. tetracycline responsive promoter, thymidine kinase promoter, SV-40 promoter and the like.

Suitable hosts for the expression of recombinant nucleic acid molecules may be prokaryotic or eukaryotic in origin. Hosts suitable for the expression of recombinant nucleic acid molecules may be selected from bacteria, yeast, insect cells and mammalian cells.

In another aspect the present invention also relates to a method of diagnosing schizophrenia and/or affective psychosis or susceptibility to schizophrenia and/or affective psychosis in an individual, wherein the method

comprises determining if the *N33* gene in the individual has been disrupted by a mutation or chromosomal rearrangement.

The methods which may be employed to elucidate such a mutation or chromosomal rearrangement are well known to those of skill in the art and could be detected for example using PCR or in hybridisation studies using suitable probes which could be designed to span an identified mutation site or chromosomal breakpoint in close proximity to the *N33* gene, such as the breakpoint identified by the present inventors and described herein.

Once a particular polymorphism or mutation has been identified it may be possible to determine a particular course of treatment. For example it is known that some forms of treatment work for some patients, but not all. This may in fact be due to mutations in the N33 gene or surrounding sequence, and it may therefore be possible to determine a treatment strategy using current therapies, based on a patient's genotype.

It will be appreciated that mutations in the gene sequence or controlling elements of a gene, eg. a promoter and/or enhance can have subtle effects such as affecting mRNA splicing/stability/activity and/or control of gene expression levels, which can also be determined. Also the relative levels of RNA can be determined using for example hybridisation or quantitative PCR as a means to determine if the *N33* gene has been disrupted.

Moreover the presence and/or levels of the *N33* gene products themselves can be assayed by immunological techniques such as radioimmunoassay, Western blotting and ELISA using specific antibodies raised against the gene products. The present invention also therefore relates to antibodies specific for a *N33* gene product and uses thereof in diagnosis and/or therapy.

A further aspect of the present invention therefore provides antibodies specific to the polypeptides of the present invention or epitopes thereof. Production and purification of antibodies specific to an antigen is a matter of ordinary skill, and the methods to be used are clear to those skilled in the art. The term antibodies can include, but is not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanised or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope binding fragments of any of the above. Such antibodies may be used in modulating the expression or activity of the particular polypeptide, or in detecting said polypeptide *in vivo* or *in vitro*.

Using the sequences disclosed herein, it is possible to identify related sequences in other animals, such as mammals, with the intention of providing an animal model for psychiatric disorders associated with the improper functioning of the nucleotide sequences and proteins of the present invention. Once identified, the homologous

sequences can be manipulated in several ways common to the skilled person in order to alter the functionality of the nucleotide sequences and proteins homologous to those of the present invention. For example, "knock-out" animals may be created, that is, the expression of the genes comprising the nucleotide sequences homologous to those of the present invention may be reduced or substantially eliminated in order to determine the effects of reducing or substantially eliminating the expression of such genes. Alternatively, animals may be created where the expression of the nucleotide sequences and proteins homologous to those of the present invention are upregulated, that is, the expression of the genes comprising the nucleotide sequences homologous to those of the present invention may be increased in order to determine the effects of increasing the expression of these genes. In addition to these manipulations, substitutions, deletions and additions may be made to the nucleotide sequences encoding the proteins homologous to those of the present invention in order to effect changes in the activity of the proteins to help elucidate the function of domains, amino acids, etc. in the proteins. Furthermore, the sequences of the present invention may also be used to transform animals to the manner described above. The manipulations described above may also be used to create an animal model of schizophrenia and/or affective psychosis associated with the improper functioning of the nucleotide sequences and/or proteins of the present invention in order to evaluate potential agents

which may be effective for combatting psychotic disorders, such as schizophrenia and/or affective psychosis.

Thus, the present invention also provides for screens for identifying agents suitable for preventing and/or treating schizophrenia and/or affective psychosis associated with disruption or alteration in the expression of the *N33* gene and/or its gene products. Such screens may easily be adapted to be used for the high throughput screening of libraries of compounds such as synthetic, natural or combinatorial compound libraries.

Thus, the *N33* gene products according to the present invention can be used for the *in vivo* or *in vitro* identification of novel ligands or analogs thereof. For this purpose binding studies can be performed with cells transformed with nucleotide fragments according to the invention or an expression vector comprising a polynucleotide fragment according to the invention, said cells expressing the *N33* gene products according to the invention.

Alternatively also the *N33* gene products according to the invention as well as ligand-binding domains thereof can be used in an assay for the identification of functional ligands or analogs for the *N33* gene products.

Methods to determine binding to expressed gene products as well as *in vitro* and *in vivo* assays to determine biological activity of gene products are well known. In general, expressed gene product is contacted with the compound to be tested and binding, stimulation or

inhibition of a functional response is measured.

Thus, the present invention provides for a method for identifying ligands for *N33* gene products, said method comprising the steps of:

a) introducing into a suitable host cell a polynucleotide fragment according to the invention;

b) culturing cells under conditions to allow expression of the polynucleotide fragment;

c) optionally isolating the expression product;

d) bringing the expression product (or the host cell from step b)) into contact with potential ligands which will possibly bind to the protein encoded by said polynucleotide fragment from step a);

e) establishing whether a ligand has bound to the expressed protein; and

f) optionally isolating and identifying the ligand.

As a preferred way of detecting the binding of the ligand to the expressed protein, also signal transduction capacity may be measured.

Compounds which activate or inhibit the function of *N33* gene products may be employed in therapeutic treatments to activate or inhibit the polypeptides of the present invention.

The present invention will now be further described by way of Example and with reference to the Figures which show:

Figure 1 shows a schematic representation of the N33 gene : exon splicing and chromosome breakpoint identified in the present invention;

Figure 2 shows the nucleotide sequence of the various exons for N33;

Figure 3 shows the various transcript options and associated amino acid sequences of the transcripts for N33;

Figure 4 shows N33 protein aligned with other homologues; and

Figure 5 shows the effect of the C-terminus of the various N33 splice forms.

Materials and methods

Lymphocyte extraction and metaphase chromosome preparation

Lymphocytes were extracted from 7mls of patient blood (for storage and generation of EBV-transformed cell lines) using density gradient separation (Histopaque-1077, Sigma). In order to generate metaphase-arrested chromosomes for cytogenetic analysis, 0.8mls of patient blood were cultured for 71hrs in medium containing phytohaemagglutinin (Peripheral Blood Medium, Sigma). The short-term cultures were treated with colcemid for one hour followed by a conventional fixing procedure. Fixed chromosomes were dropped onto microscope slides and stored for 1 week prior to use in FISH experiments.

Selection of YAC clones for FISH probe synthesis

YAC clones were selected from the Whitehead/MIT map of the relevant chromosome in the cytogenetic intervals within which the breakpoints were adjudged to lie. YACs were obtained from the HGMP Resource Centre, Babraham Bioincubator, Babraham, Cambridge, UK (http://www.hgmp.mrc.ac.uk/). Clone DNA was prepared by standard methods and PCR amplified using primers designed against consensus sequence elements within the archetypal *Alu* repeat, Breen et al, 1992. This "Alu-PCR" gives a representative spread of non-repetitive sequence over the full length of the YAC and generates a better FISH probe than native YAC DNA. Alu-PCR was performed using the Expand Long Template PCR kit (Roche). Cycling conditions:94°C - 45s, 55°C - 30s, 68°C - 8min: 35 cycles. 68°C - 10min final extension.

Fluorescence *in situ* hybridisation (FISH) protocol

Probe template DNA (pooled Alu-PCR products, BAC clone DNA, cosmid clone DNA or long-range PCR products) was labelled by nick translation and hybridised to patient metaphase spreads using standard FISH methods. Slides were counterstained with DAPI in Vectashield anti-fade solution (Vector laboratories). The Zeiss Axioskop fluorescence microscope with a chroma number 81000 multi-spectral filter set. Images were captured using Vysis SmartCapture extension running within IP Lab spectrum. FISH signals observed on derived chromosomes dictated the selection of

further clones required to "walk" towards the breakpoint. Breakpoint-spanning FISH probes have signals on a normal chromosome and on both derived chromosomes.

Resolution of breakpoint position

BAC clones corresponding to positive YAC regions were arranged into contigs by consulting the Washington University FPC (http://www.genome.wustl.edu/gsc/human/Mapping/index.shtml), UCSC GoldenPath Draft Human Genome Browser (http://genome.ucsc.edu/goldenPath/hgTracks.html) and Ensembl (http://www.ensembl.org/) databases. BAC clones were supplied by BACPAC Resources, Oakland, California, USA (http://www.chori.org/bacpac/). Clone selection was biased to gene-containing BACs. Once a breakpoint-spanning BAC was identified, the position of the breakpoint in relation to candidate gene exons was determined by FISH probes generated from chromosome-specific library cosmids (HGMP Resource centre) or precisely positioned, repeat element-free long-range PCR products (Expand long range PCR kit, Roche; see below for primer sequences). Cycling conditions:94°C - 45s, 52°C - 30s, 68°C - 11min: 35 cycles. 68°C - 15min final extension. Cosmids were isolated by probing the appropriate chromosome-specific library filters (HGMP-RC) with isotopically labelled exon-specific PCR products.

In this case, primers corresponding to N33 3'UTR sequences and an STS, SHGC-12093 (Acc. No. G17275) were designed (see below for primer sequences). These PCR products were used to screen the chromosome 8 specific cosmid library (LA08). Among others, positive cosmids LA0854-H5 (3' UTR) and LA08145-E3 (STS) were isolated and subsequently used in FISH experiments (see below for results).

3'UTR primers
Primer A: TGCCACGTGTTAGCAGAAAG
Primer B: TGCCTTTAACCAGATGAGGC

SHGC-12093 primers
Primer A: TCTTGTGGGTCACAATTAGGC
Primer B: TAAAAAGGTGCAGTTTCTTCAGC'.

Example 1: Molecular characterisation of chromosomal disruption and identification of disrupted gene

The subject has schizoaffective disorder and a balanced reciprocal translocation between chromosomes 3 and 8. Chromosome 3 sequences were not examined in any detail because the breakpoint lies within a region of the genome for which no good sequence contig is available. However, RPCI-11 BACs 268j14 (acc. no. AC022848) and 508o18 (acc. no. AC073518) lie either side of the 3p12 breakpoint region. A 8p22 breakpoint-crossing YAC, 931_a_1, was identified. This permitted a 8p22 breakpoint -crossing BAC

RPCI-11 23j14 (acc. no. AC019292) to be found. This was shown to contain the 3' end of the N33 gene (FIG.2). Subsequently, FISH with cosmids LA0854-H5 and LA08145-E3 from the LANL chromosome 8 specific library (HGMP Resource Centre, Babraham, Cambridge, UK) flanked the breakpoint, placing it approximately 100Kb from exon 11 of N33. N33 is related to a number of genes, human IAG2, Drosophila CG7830, C. elegans g304348 and two yeast proteins, OST3 and OST6 (see fig 4 for alignment of proteins). While the homologies between N33 and the yeast proteins are relatively weak, they share conserved cysteine residues and have the same locations for the four transmembrane domains as predicted by hydropathy plots. Ost3 and Ost6 are components of the oligosaccharyl transferase complex responsible for the addition of oligosaccharides to selected proteins. This has been backed up by protein structure prediction programs detailed in a recent report Fetrow et al, 2001.

The present inventors have identified an alternative start exon, herein identified as exon 1a (see Figures 1 & 2) to that in the public database, herein identified at exon 1b. Additionally they have identified a complex variation of splicing with the exons and proposed sequences of the transcripts, shown in Figures 1, 2 and 3 respectively. In view of the complex splice variations the C-terminal sequence of the various N3 splice forms is predicted to vary and this is shown in Figure 5.

Because N33 lies within a linkage hotspot for

Because N33 lies within a linkage hotspot for schizophrenia (Gurling et al, 2001, Brzustowicz et al, 1999, Blouin et al, 1998, Kaufmann et al, 1998, Kendler et al, 1996, Pulver et al, 1995) the present inventors decided to carry out an association study on this gene. Three microsatellite markers (D8S549, N33 microsatellite and D8S1992

Microsatellites used in associated study

D8S549

Primer A: AAATGAATCTCTGATTAGCCAAC

Primer B: TGAGAGCCAACCTATTTCTACC

N33 microsatellite

Primer A: AGGCTGAGTGCCAAAAAGTA

Primer B: CTTTAAGCTTGCTATTTGAAGGC

D8S1992

Primer A: TTCATCGTCTGAACCTGG

Primer B: ACACATTTCCTCTATGTTGC) were chosen and used to type 25 mother-father-schizophrenic proband trios and 64 schizophrenic cases and 64 normal controls. The haplotypes derived from the trio study were examined for frequency bias in the case and control samples. Certain haplotypes are currently over-represented in the schizophrenic case genotypes compared to controls. Appropriate individuals with the haplotypes are currently being screened for mutations.

REFERENCES

1. Berrettini W.H. (2000) Genetics of psychiatric disease. Annu. Rev. Med. 51:465 - 79.

2. Lai C.S, Fisher S.E, Hurst J.A, Vargha-Khadem F, Monaco A.P. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 413(6855):519 - 23.

3. Kleinjan D.J, van Heyningen V. (1998) Position effect in human genetic disease. Hum Mol Genet. 7(10):1611 - 8.

4. Brennan J, Skarnes W.C. (1999) Gene trapping in mouse embryonic stem cells. Methods Mol. Biol. 97:123 - 38.

5. Breen M, Arveiler B, Murray I, Gosden J.R, Porteous D.J. (1992) YAC mapping by FISH using Alu-PCR-generated probes. Genomics. 13(3):726 - 30.

6. Fetrow J.S, Siew N, Di Gennaro J.A, Martinez-Yamout M, Dyson H.J, Skolnick J. (2001) Genomic-scale comparison of sequence- and structure-based methods of function prediction: does structure provide additional insight? Protein Sci. 2001 May;10(5):1005 - 14.

7. Gurling H.M, Kalsi G, Brynjolfson J, Sigmundsson T, Sherrington R, Mankoo B.S, Read T, Murphy P, Blaveri E, McQuillin A, Petursson H, Curtis D. (2001) Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23. Am J Hum Genet. Mar;68(3):661 - 73.

8. Brzustowicz L.M, Honer W.G, Chow E.W, Little D, Hogan J, Hodgkinson K, Bassett AS. (1999) Linkage of familial schizophrenia to chromosome 13q32. Am J Hum Genet. Oct;65(4):1096 - 103.

9. Blouin J.L, Dombroski B.A, Nath S.K, Lasseter V.K, Wolyniec P.S, Nestadt G, Thornquist M, Ullrich G, McGrath J, Kasch L, Lamacz M, Thomas M.G, Gehrig C, Radhakrishna U, Snyder S.E, Balk K.G, Neufeld K, Swartz K.L, DeMarchi N, Papadimitriou G.N, Dikeos D.G, Stefanis C.N, Chakravarti A, Childs B, Pulver A.E, et al. (1998) Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. Nat Genet. Sep;20(1):70 - 3.

10. Kaufmann C.A, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel P.D, Meyer J, Zambuto C.T, Schmitt K, Matise T.C, Harkavy Friedman J.M, Hampe C, Lee H, Shore D, Wynne D, Faraone S.V, Tsuang M.T, Cloninger C.R. (1998) NIMH Genetics Initiative Millenium Schizophrenia Consortium: linkage analysis of African-American pedigrees. Am J Med Genet. Jul 10;81(4):282 - 9.

11. Kendler K.S, MacLean C.J, O'Neill F.A, Burke J, Murphy B, Duke F, Shinkwin R, Easter S.M, Webb B.T, Zhang J, Walsh D, Straub (1996) Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish Study of High-Density Schizophrenia Families. Am J Psychiatry. Dec;153(12):1534 - 40.

12. Pulver A.E, Lasseter V.K, Kasch L, Wolyniec P, Nestadt G, Blouin J.L, Kimberland M, Babb R, Vourlis S, Chen H, et al. (1995) Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. Am J Med Genet. Jun 19;60(3):252 - 60.

13. MacGrogan D, Levy A, Bova G.S, Isaacs W.B and Bookstein R. Structure and methylation-associated silencing of a gene with a homozygously deleted region of himan chromosome band 8p22. Genomics 35 (1), 55 - 65 (1996).

Figure 1

N33 gene: exon splicing and chromosome breakpoint

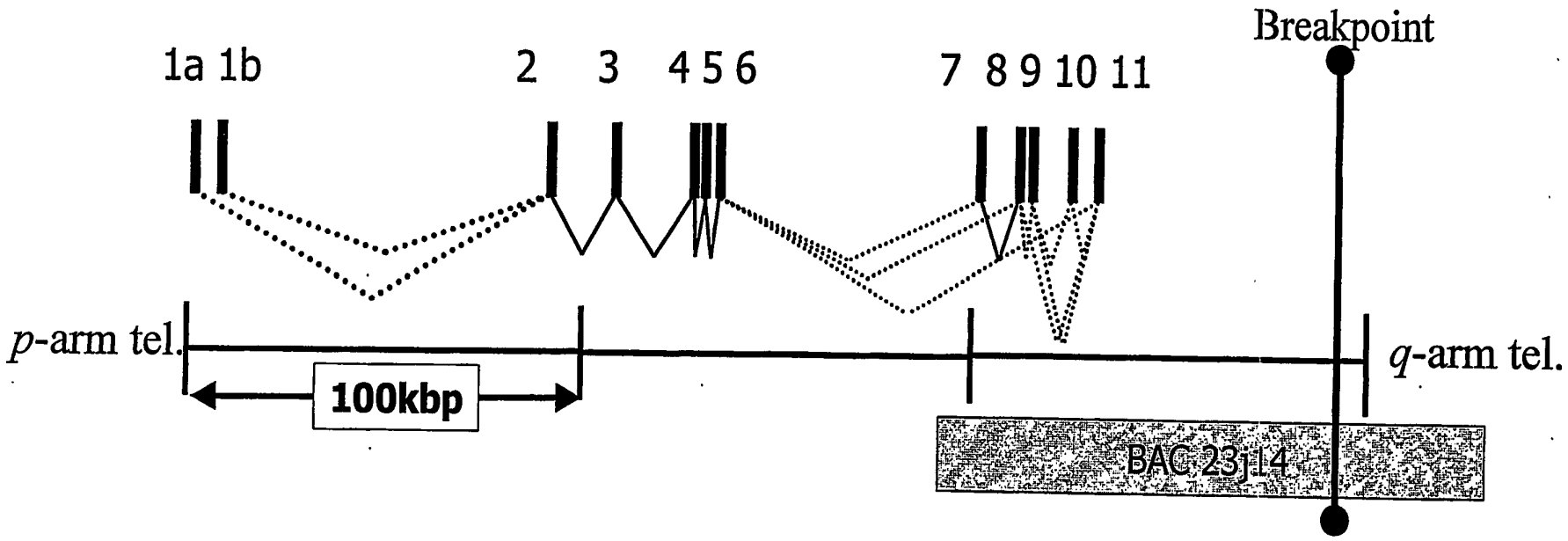


Figure 2

1a

atcttcctcctgctctggctgtgtgaagatctgcctccttcctcttcggcttcatgcatgatcgtaagtttcctga
ggcctcctcagccatgcttcctgcatagcctgcagaaat

1b

cccgggtccctcgcaaagccgctgccatcccggagggcccagccagcgggctcccggaggctggccgggcaggcgt
ggtgcgcggtaggagctgggcgcgcacggctaccgcgcgtggaggagacactgccctgccgcgatgggggcccggg
gcgctccttcacgccgtaggcaagcggggcggcggctgcggtacctgcccaccgggagctttcccttccttctcct
gctgctgctgctctgcatccagctcgggggaggacagaagaaaaaggag

2-6
These exons have been joined together as they are always spliced in this way.
aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatcttccgaatgaatggtgata
aattccgaaaatttataaaggcaccacctcgaaactattccatgattgttatgttcactgctcttcagcctcagcg
gcagtgttctgtgtgcaggcaagctaatgaagaatatcaaatactggcgaactcctggcgctattcatctgctttt
tgtaacaagctcttcttcagtatggtggactatgatgaggggacagacgtttttcagcagctcaacatgaactctg
ctcctacattcatgcattttcctccaaaaggcagacctaagagagctgatacttttgacctccaaagaattggatt
tgcagctgagcaactagcaaagtggattgctgacagaacggatgttcatattcgggttttcagaccacccaactac
tctggtaccattgctttggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggagt
tcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgacttctggccagatgtggaa
ccatatccgtggacctccatatgctcataagaacccacacaatggacaagtg

seven

agctacattcatgggagcagccaggctcagtttgtggcagaatcacacattattctggtactga

eight

atgccgctatcaccatggggatggttcttctaaatgaagcagcaacttcgaaaggcgatgttggaaaaagacgga

eight+
This is identical to 8 except a cryptic splice acceptor upstream is employed

tttaaccattctggaacattgtgttcagagccagaaaaattaatagattttattcacatctatgtctacggcttcc
ttgacaactactgcagatgccgctatcaccatggggatggttcttctaaatgaagcagcaacttcgaaaggcgatg
ttggaaaaagacgga

nine

taatttgcctagtgggattgggcctggtggtcttcttcttcagttttctactttcaatatttcgttccaagtacca
cggctatccttatag

ten

tgatctggactttgagtgagaagatgtgatttggaccatggcacttaaaaactctataacctcag

eleven

ctttttaattaaatgaagccaagtgggatttgcataaagtgaatgtttaccatgaagataaactgttcctgacttt
atactattttgaattc

Figure 3

Alternative start exons

If 1a is used then protein will begin:


..... (encoded by exon 2).

If 1b is used (better KOZAK consensus) then protein will have an additional 56 amino acids on the N-terminus:

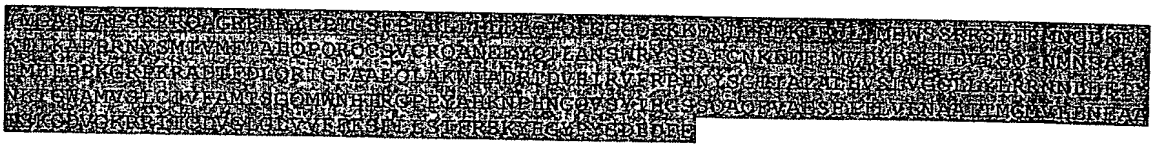

.....

Transcript options

Below are the splicing combinations of the N33 gene observed to date in the public EST database. Apart from the alternative start exons, there are different combinations of exons after exon 6. Green is the nucleic acid sequence, Blue/Yellow is the resulting amino acid sequence.

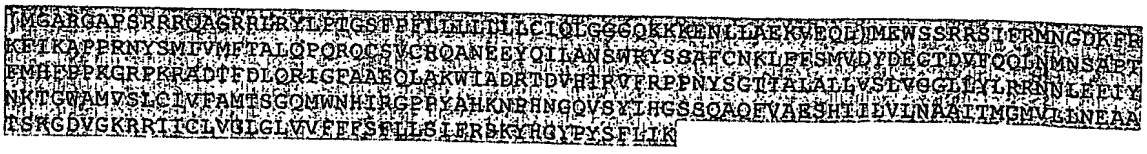
2-6,7,8,9,10,11

aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatcttccgaatgaatggtgata
aattccgaaaatttataaaggcaccacctcgaaactattccatgattgttatgttcactgctcttcagcctcagcg
gcagtgttctgtgtgcaggcaagctaatgaagaatatcaaatactggcgaactcctggcgctattcatctgctttt
tgtaacaagctcttcttcagtatggtggactatgatgaggggacagacgtttttcagcagctcaacatgaactctg
ctcctacattcatgcattttcctccaaaaggcagacctaagagagctgatacttttgacctccaaagaattggatt
tgcagctgagcaactagcaaagtggattgctgacagaacggatgttcatattcgggttttcagaccacccaactac
tctggtaccattgctttggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggagt
tcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgacttctggccagatgtggaa
ccatatccgtggacctccatatgctcataagaacccacacaatggacaagtgagctacattcatgggagcagccag
gctcagtttgtggcagaatcacacattattctggtactgaatgccgctatcaccatggggatggttcttctaaatg
aagcagcaacttcgaaaggcgatgttggaaaaagacggataatttgcctagtgggattgggcctggtggtcttctt
cttcagttttctactttcaatatttcgttccaagtaccacggctatccttatagtgatctggactttgagtgagaa
gatgtgatttggaccatggcacttaaaaactctataacctcagcttttaattaaatgaagccaagtgggatttgc
ataaagtgaatgtttaccatgaagataaactgttcctgactttatactattttgaattc



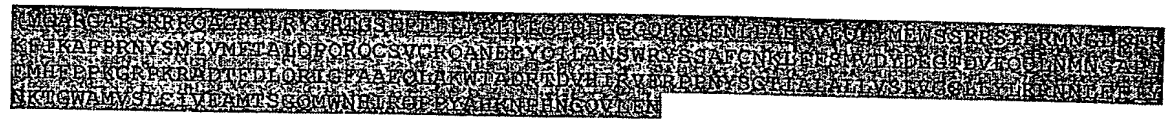
2-6,7,8,9,11

aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatcttccgaatgaatggtgata
aattccgaaaatttataaaggcaccacctcgaaactattccatgattgttatgttcactgctcttcagcctcagcg
gcagtgttctgtgtgcaggcaagctaatgaagaatatcaaatactggcgaactcctggcgctattcatctgctttt
tgtaacaagctcttcttcagtatggtggactatgatgaggggacagacgtttttcagcagctcaacatgaactctg
ctcctacattcatgcattttcctccaaaaggcagacctaagagagctgatacttttgacctccaaagaattggatt
tgcagctgagcaactagcaaagtggattgctgacagaacggatgttcatattcgggttttcagaccacccaactac
tctggtaccattgctttggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggagt
tcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgacttctggccagatgtggaa
ccatatccgtggacctccatatgctcataagaacccacacaatggacaagtgagctacattcatgggagcagccag
gctcagtttgtggcagaatcacacattattctggtactgaatgccgctatcaccatggggatggttcttctaaatg
aagcagcaacttcgaaaggcgatgttggaaaaagacggataatttgcctagtgggattgggcctggtggtcttctt
cttcagttttctactttcaatatttcgttccaagtaccacggctatccttatagcttttaattaaatgaagccaa
gtgggatttgcataaagtgaatgtttaccatgaagataaactgttcctgactttatactattttgaattc



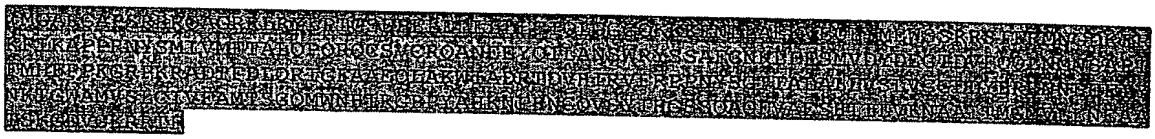
2-6,11

aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatcttccgaatgaatggtgata
aattccgaaaatttataaaggcaccacctcgaaactattccatgattgttatgttcactgctcttcagcctcagcg
gcagtgttctgtgtgcaggcaagctaatgaagaatatcaaatactggcgaactcctggcgctattcatctgctttt
tgtaacaagctcttcttcagtatggtggactatgatgaggggacagacgtttttcagcagctcaacatgaactctg
ctcctacattcatgcattttcctccaaaaggcagacctaagagagctgatacttttgacctccaaagaattggatt
tgcagctgagcaactagcaaagtggattgctgacagaacggatgttcatattcgggttttcagaccacccaactac
tctggtaccattgctttggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggagt
tcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgacttctggccagatgtggaa
ccatatccgtggacctccatatgctcataagaacccacacaatggacaagtgcttttaattaaatgaagccaagt
gggatttgcataaagtgaatgtttaccatgaagataaactgttcctgactttatactattttgaattc



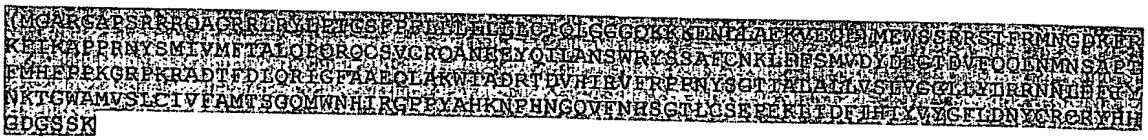
2-6,7,8,11

aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatcttccgaatgaatggtgata
aattccgaaaatttataaaggcaccacctcgaaactattccatgattgttatgttcactgctcttcagcctcagcg
gcagtgttctgtgtgcaggcaagctaatgaagaatatcaaatactggcgaactcctggcgctattcatctgctttt
tgtaacaagctcttcttcagtatggtggactatgatgaggggacagacgtttttcagcagctcaacatgaactctg
ctcctacattcatgcattttcctccaaaaggcagacctaagagagctgatacttttgacctccaaagaattggatt
tgcagctgagcaactagcaaagtggattgctgacagaacggatgttcatattcgggttttcagaccacccaactac
tctggtaccattgctttggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggagt
tcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgacttctggccagatgtggaa
ccatatccgtggacctccatatgctcataagaacccacacaatggacaagtgagctacattcatgggagcagccag
gctcagtttgtggcagaatcacacattattctggtactgaatgccgctatcaccatggggatggttcttctaaatg
aagcagcaacttcgaaaggcgatgttggaaaaagacggacttttaattaaatgaagccaagtgggatttgcataa
agtgaatgtttaccatgaagataaactgttcctgactttatactattttgaattc



2-6,8+,9,11

aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatcttccgaatgaatggtgata
aattccgaaaatttataaaggcaccacctcgaaactattccatgattgttatgttcactgctcttcagcctcagcg
gcagtgttctgtgtgcaggcaagctaatgaagaatatcaaatactggcgaactcctggcgctattcatctgctttt
tgtaacaagctcttcttcagtatggtggactatgatgaggggacagacgtttttcagcagctcaacatgaactctg
ctcctacattcatgcattttcctccaaaaggcagacctaagagagctgatacttttgacctccaaagaattggatt
tgcagctgagcaactagcaaagtggattgctgacagaacggatgttcatattcgggttttcagaccacccaactac
tctggtaccattgctttggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggagt
tcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgacttctggccagatgtggaa
ccatatccgtggacctccatatgctcataagaacccacacaatggacaagtgtttaaccattctggaacattgtgt
tcagagccagaaaaattaatagattttattcacatctatgtctacggcttccttgacaactactgcagatgccgct
atcaccatggggatggttcttctaaatgaagcagcaacttcgaaaggcgatgttggaaaaagacggataatttgcc
tagtgggattgggcctggtggtcttcttcttcagttttctactttcaatatttcgttccaagtaccacggctatcc
ttatagcttttaattaaatgaagccaagtgggatttgcataaagtgaatgtttaccatgaagataaactgttcct
gactttatactattttgaattc



2-6,8+,11

aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatcttccgaatgaatggtgata
aattccgaaaatttataaaggcaccacctcgaaactattccatgattgttatgttcactgctcttcagcctcagcg
gcagtgttctgtgtgcaggcaagctaatgaagaatatcaaatactggcgaactcctggcgctattcatctgctttt
tgtaacaagctcttcttcagtatggtggactatgatgaggggacagacgtttttcagcagctcaacatgaactctg
ctcctacattcatgcattttcctccaaaaggcagacctaagagagctgatacttttgacctccaaagaattggatt
tgcagctgagcaactagcaaagtggattgctgacagaacggatgttcatattcgggttttcagaccacccaactac
tctggtaccattgctttggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggagt

tcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgacttctggccagatgtggaa
ccatatccgtggacctccatatgctcataagaacccacacaatggacaagtgtttaaccattctggaacattgtgt
tcagagccagaaaaattaatagattttattcacatctatgtctacggcttccttgacaactactgcagatgccgct
atcaccatggggatggttcttctaaatgaagcagcaacttcgaaaggcgatgttggaaaaagacggacttttaat
taaatgaagccaagtgggatttgcataaagtgaatgtttaccatgaagataaactgttcctgactttatactattt
tgaattc

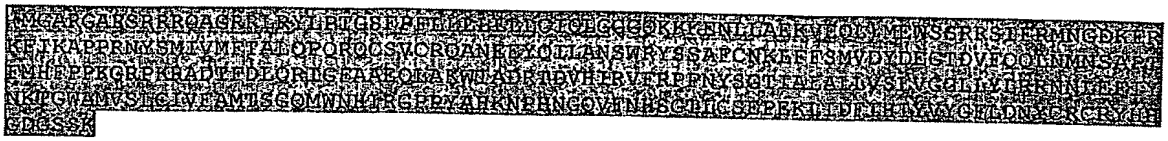


Figure 4 N33 protein: alignment with homologs

```
IAG2_HUMAN          MAAR-----------WRFWCVSVTMVVALLIVCDVPSASAQRKKE-MVLSEKVSQLMEWTNKRPVIRMNGDKFRRLVKAPP
N33_HUMAN           MGARGAPSRRRQAGRRLRYLPTGSFPFLLLLLLLCIQLGGGQKKKE-NLLAEKVEQLMEWSSRRSIFRMNGDKFRKFIKAPP
DROS._CG7830        ---------------MRLLHKTLLSGLLVVALFAIYAAAQSKSKTGLSLSEKVQNLVDMNAKKPLLRFNGPKFREYVKSAP
Celegans_g304348    ------------------------------MLLAVYESAQQQT-----LEDKVQNLVDLTSRQSIVKFNMDKWKTLVRMQP
Yeast_Ost3p         -------------------------MNWLFLVSLVFFCGVSTHPALAMSSNRLLKLANKSPKK---IIPLKDSSFENILAP
Yeast_Ost6p         ----------------------MKWCSTYIIIWLAIIFHKFQKSTA--TASHNIDDILQLKDDTGVITVTADNYPLLSRGVP
                                                                                                    -

IAG2_HUMAN          --RNYSVIVMFTALQLHRQCVVCKQADEEFQILANSWRYSSAFTN------RIFFAMVDFDEG----SDVFQMLNMNSAPTF
N33_HUMAN           --RNYSMIVMFTALQPQRQCSVCRQANEEYQILANSWRYSSAFCN------KLFFSMVDYDEG----TDVFQQLNMNSAPTF
DROS._CG7830        --RNYSMIVMLTALAPSRQCQICRHAHDEFAIVANSYRFSSTYSN------KLFFAMVDFDDG----SEVFQLLRLNTAPVF
Celegans_g304348    --RNYSMIVMFTALSPGVQCPICKPAYDEFMIVANSHRYTSSEGDRR----KVFFGIVDYEDA----PQIFQQMNLNTAPIL
Yeast_Ost3p         PHENAYIVALFTATAPEIGCSLCLELESEYDTIVASWFDDHPDAKSSNSDTSIFFTKVNLEDPSKTIPKAFQFFQLNNVPRL
Yeast_Ost6p         GYFNILYITMRGTNSNGMSCQLCHDFEKTYHAVADVIRSQAPQSLN-------LFFTVDVNEV----PQLVKDLKLQNVPHL
                                          CYST.

IAG2_HUMAN          INFPAK-GKPKRGDTYELQVRG--FSAEQIARWIADR-----TDVNIRVIRPPNYAGPLMLGLLLAVIGGLVYLRRSNMEF-
N33_HUMAN           MHFPPK-GRPKRADTFDLQRIG--FAAEQLAKWIADR-----TDVHIRVFRPPNYSGTIALALLVSLVGGLLYLRRNNLEF-
DROS._CG7830        MHFPAK-GKPKGADTMDIHRVG--FAADSIAKFVAER-----TDITIRIFRPPNYSGTVAMITLVALVGSFLYIRRNNLEF-
Celegans_g304348    YHFGPKLGAKKRPEQMDFQRQG--FDADAIGRFVADQ-----TEVHVRVIRPPNYTAPVVIALFVALLLGMLYMKRNSLDF-
Yeast_Ost3p         FIFKPNSPSILDHSVISISTDTGSERMKQIIQAIKQF-----SQVNDFSLHLPMDWTPIITSTIITFITVLLFKKQSKLMFS
Yeast_Ost6p         VVYPPAESNKQSQFEWKTSPFYQYSLVPENAENTLQFGDFLAKILNISITVPQAFNVQEFVYYFVACMVVFIFIKKVILPKV
                                                                                ******TM 1*******ccccccccccc

IAG2_HUMAN          -LFNKTGWAFAALCFVLAMTSGQMWNHIRGPPYAHKNPHTGHVNYIHGSSQAQFVAETHIVLLFNGGVTLGMVLLCEAATSD
N33_HUMAN           -IYNKTGWAMVSLCIVFAMTSGQMWNHIRGPPYAHKNPHNGQVSYIHGSSQAQFVAESHIILVLNAAITMGMVLLNEAATSK
DROS._CG7830        -LYNKNLWGAIAVFFCFAMISGQMWNHIRGPPLVHKS-QNGGVAYIHGSSQGQLVVETYIVMFLNAMIVLGMILLIESGTPK
Celegans_g304348    -LFNRTVWGFVCLAITFIFMSGQMWNHIRGPPFMITNPNTKEPSFIHGSTQFQLIAETYIVGLLYALIAIGFICVNEAADQS
Yeast_Ost3p         IISSRIIWATLSTFFIICMISAYMFNQIRNTQLAGVGPKGEVMYFLPNEFQHQFAIETQVMVLIYGTLAALVVVLVKGIQFL
Yeast_Ost6p         TNKWKLFSMILSLGILLPSITGYKFVEMNAIPFIARDAKN-RIMYFSGGSGWQFGIEIFSVSLMYIVMSALSVLLIYVPKIS
                    ccccc*******TM 2******                                          ******TM 3******ccccccccc

IAG2_HUMAN          MDIGKR----------------KIMCVAGIGLVVLFFSWMLSIFRSKYHGYPYSFLMS----------------
N33_HUMAN           GDVGKR----------------RIICLVGLGLVVFFFSFLLSIFRSKYHGYPYSDLDFE-(1)------------
DROS._CG7830        AHN-KN----------------RIMAMTGLVLLTVFFSFLLSVFRSKAQGYPYISCSNRIDCSPVPVQVHPISFL
Celegans_g304348    NSKDRKNAGKKLNPLSLLNIPTNTLAIAGLVCICVFFSFLLSVFRSKYRGYPYSFLFA----------------
Yeast_Ost3p         RSHLYP-------ETKKAYFIDAILASFCALFIYVFFAALTTVFTIKSPAYPFPLLRLSAPFK------------
Yeast_Ost6p         CVSEKMR----------------GLLSSFLACVLFYFFSYFISCYLIKNPGYPIVF--------------------
                    ccccccccccccccccccccccc*****TM 4*****
                                          TF (3)                                FLIK (2)
```

Figure 5

C-termini of N33 splice forms

As mentioned above, the variety of splice forms at the 3' end of the gene has implications for the C-terminus of the protein. This is especially important when it is considered that N33 is likely to reside in the Golgi/ER compartment of the cell where C-termini are often involved in anchoring or trafficking proteins to different organelles. The blue shading indicates putative transmembrane domains. Hence, only the spliceforms with exons 1a/1b,2-6,7,8,9,10,11 or 1a/1b,2-6,7,8,9,11 are likely to encode functional proteins and these will only differ in the extreme C-terminal residues.